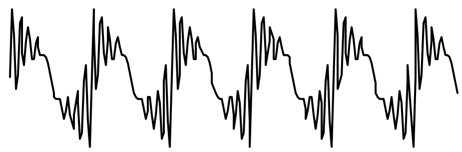
Networked Multimedia: Sound



Course Outline (2)

3. networked multimedia (sessions 10-15)
    a. sound
    b. graphics
    c. video
    d. priority, rate control, flow control
    e. middleware

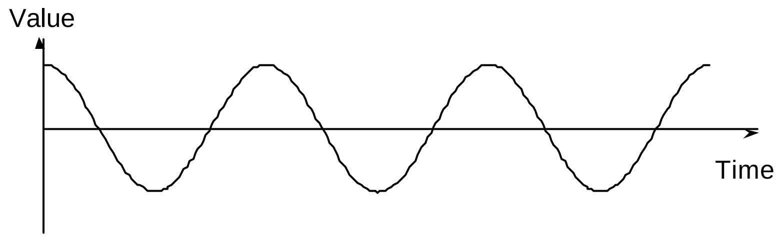
MV4924 Projects

2. (week 5) stream a sound file to an audio player

Students who have already done projects similar to 1 thru 4 may substitute a different project, in consultation with the instructor.

Networked Multimedia: Sound

Analog Signals: Sine Wave

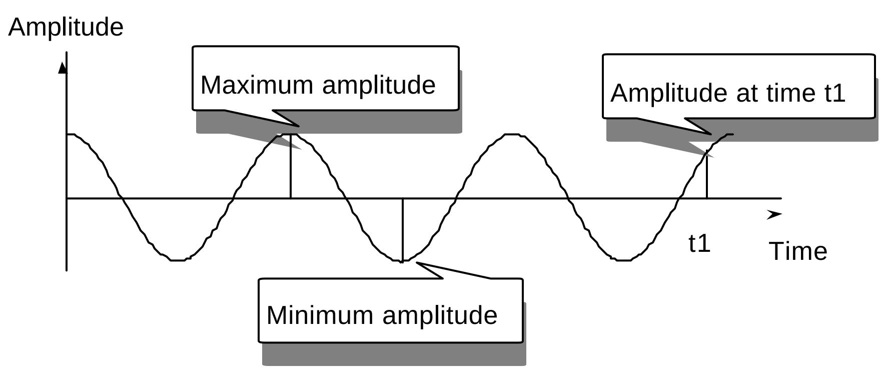


A ***sine wave*** can be fully described by three characteristics:

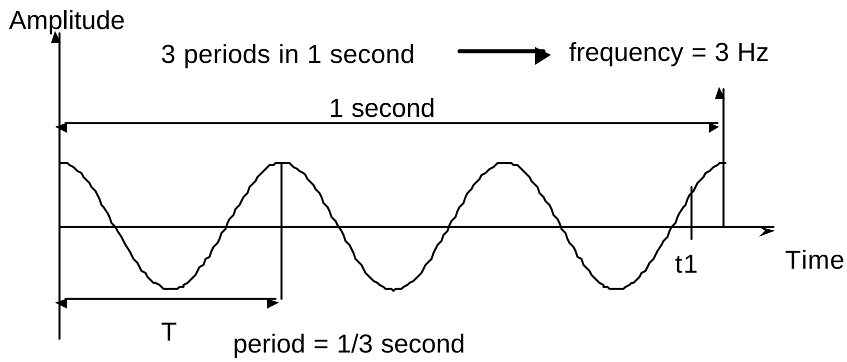
amplitude, *period* or *frequency*, and *phase*

A sine wave is the simplest analog signal; it cannot be decomposed into simpler signals.

Amplitude



Period and Frequency

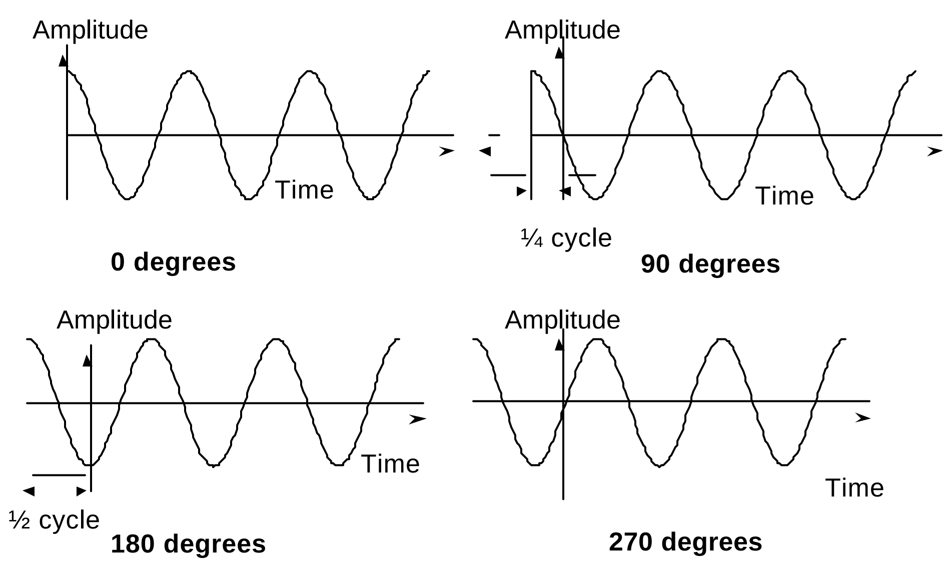


Frequency *(f)* refers to number of *periods* (*T*) in 1 second

$$T = 1/f$$

Change in a short span of time means high frequency; otherwise, means low frequency.

Phase of Cosine Waves



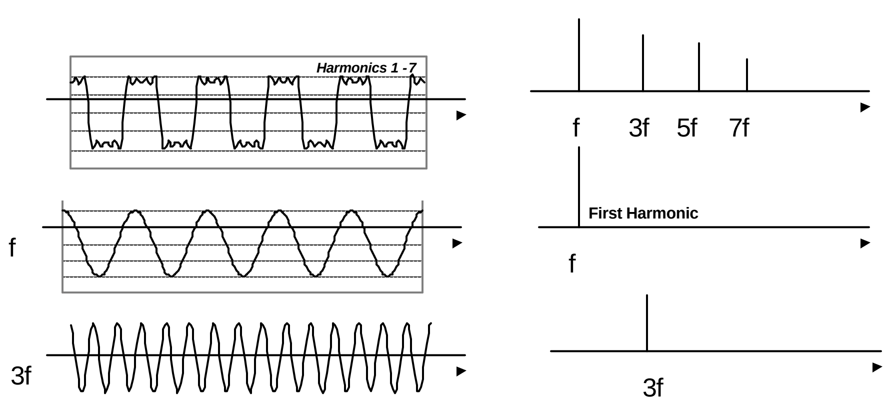
Fourier Analysis

- Consider a periodic function *g(t).*
  - f_0 is the fundamental frequency of periodic change of the signal's pattern
  - *d* is the DC component
- Then the signal can be described to any level of accuracy by the Fourier Series

$$g(t) = d + \sum_{n=1} a_n \sin(2\pi n f_o t - \phi_n \pi / 180)$$

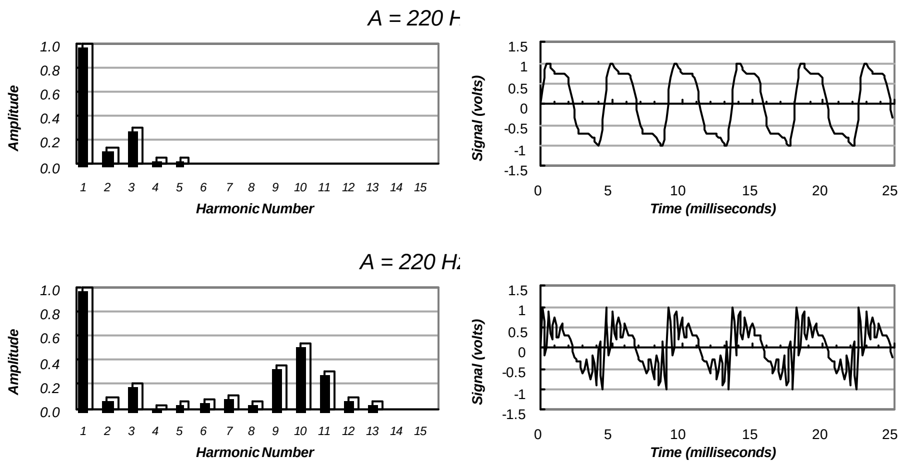
- In other words, a repetitive signal of any shape can be described as the sum of sinusoidal signals plus a constant (DC) component
- Generally, the values of a_n grow smaller as n increases; so the series up to the point where a_n is negligible is a good approximation of g(t)

Composite Waveform

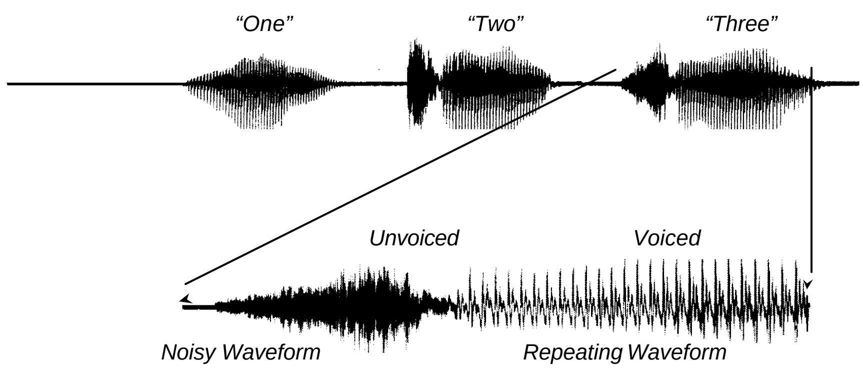


Time domain **Frequency domain**

Sound Spectra

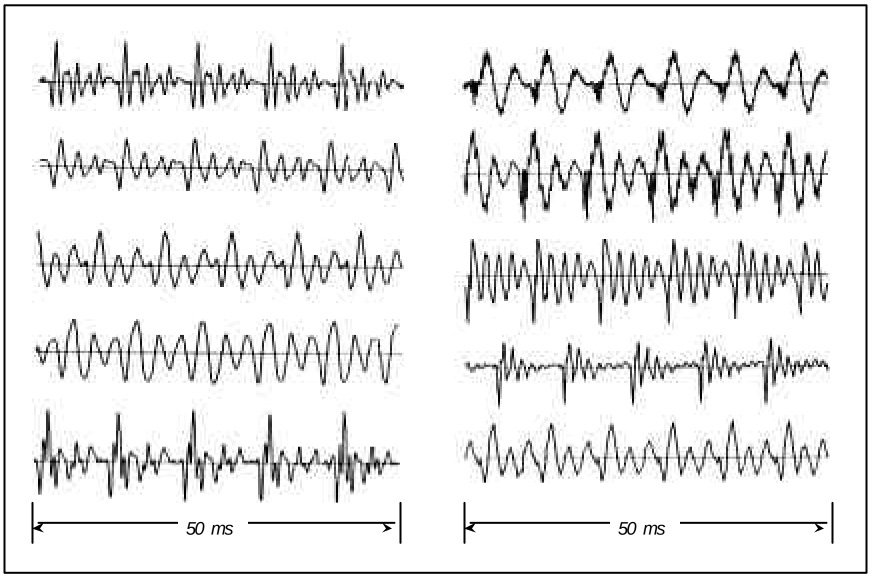


Speech Signals

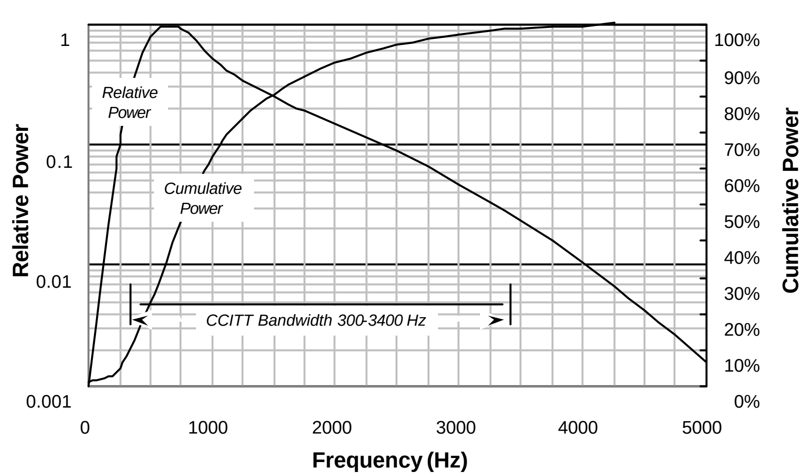


- Voiced speech is characterized by periodic vibrations of the vocal chords
  - Lungs plus vocal chords provide puffs of air
  - Speech waveform shaped by elements of the vocal tract
- Unvoiced speech characterized by noise-like waveform
  - Vocal chords unmoving
  - Speech waveform formed by air forced past lips/teeth/tongue/palate

Vowel Waveforms



Sound Power in Speech



Analog-to-Digital Conversion

- Sample analog voice signal at the Nyquist rate = 2 f_H (twice the highest frequency if $f_L = 0$), or
  - 2 X 4000 Hz = 8000 samples per second
- Convert each sample to an 8-bit binary number (called quantizing) using Pulse Code Modulation (PCM)
- Send or store this digital data as
  - 8 (bit samples) X 8000 (samples per second), or 64,000 bps
- If transmission or storage is scarce, use compression
  - trade processing for network capacity/disk space

Fidelity of Digital Sound

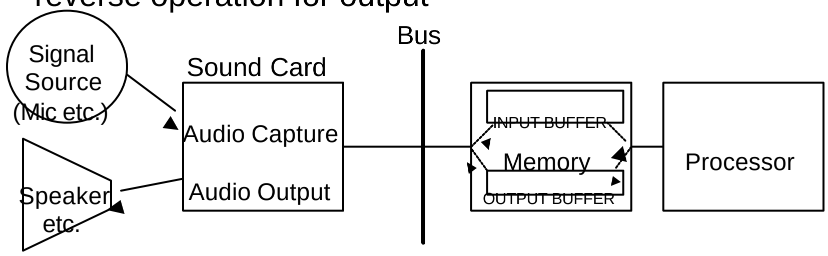
- Human ear is capable of frequency range:
  - 20 - 20,000 Hz
- And dynamic range
  - 10^{14} = 14 Bels = 140 deciBels (dB)
- We can approximate as:
  - 8,000 sample/sec = 4 kHz (toll quality telephone)
    - also 5.5kHz, 11kHz, 22kHz (CD quality)
  - 48.2dB from 8 bits
  - 96.3 dB from 16 bits

Audio Compression

- Differential PCM - encodes each signal as a change from the previous one; allows a reduced voice digitization rate (VDR)
- Delta Modulation - Similar to Differential PCM but sends the shape of the signal instead of the level (further reduces VDR)
- Linear Predictive Coding (LPC) - represents speech as a progression of phonemes (component sounds) to achieve VDRs as low as 2.4 kbps
- Common Internet phone formats:
  - Fraunhofer GSM 6.10 Groupe Spécial Mobile (European cell phone)
  - ADPCM Microsoft differential PCM
  - LPC and LPC-10 see above

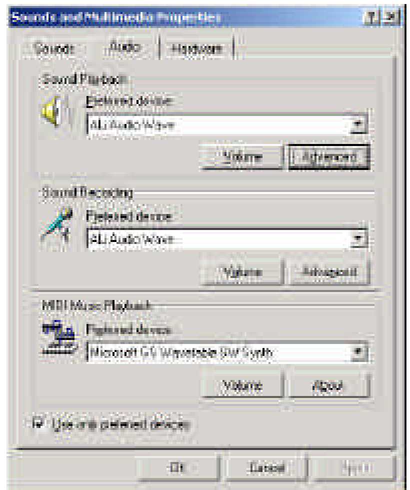
Computer Sound Architecture

- Sound capture is through a separate A-D system
  - external signal, amplified, digitized in a coder/decoder (codec)
  - PCM version stored in memory
  - processor interrupt when sample ready to compress or otherwise process
  - reverse operation for output

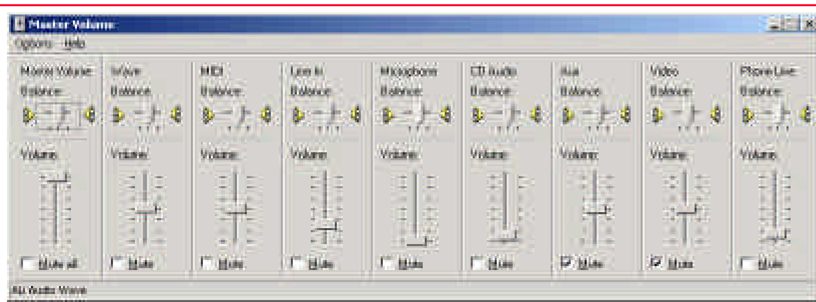


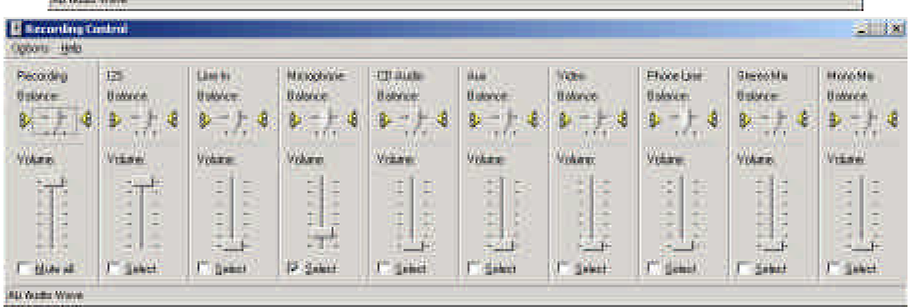
Typical Sound Options

- settings.control panel.sounds&multimedia



Sound Mixer Controls





Audio Communication Software Operation

- Speak Freely audio:
  - Main: control thread, message pump
  - Init: reads configuration files, starts listening
  - Frame: control for all options, audio listen interface, network receive function
  - Connect: network send function, connection profiles, audio talk interface, sound files
  - Dialog: popup boxes, including VoxMonitor modified extensively by NETLAB
  - Many more: ANSWER, compression, CRC, DESKEY, FACE, G711, LOOPBACK, LWL, ULAW, UTILITY...
  - NETLAB added: Mixer (with ACK to UCL), Tunnel

Common Audio File Formats

- Pure audio:
  - .wav Windows formatted sound
  - .au Sun audio
- Multimedia:
  - .avi Windows audiovisual
  - .mov Apple QuickTime Movie
- Audio recordings
  - .mp3 MPEG "layer 3" audio

Audio for NVE

- Sound effects
  - typically a repeating “loop”
  - motors, gunfire, waves, fire crackling...
  - play a sample over and over
- Simulate radio or telephone
  - basically the same as net telephone
  - for more fidelity:
    - – more bits per sample (16 bits common)
    - – more samples/second (22 k samples/sec = 11 kHz)